

Under load: The effect of verbal and motoric cognitive load on gesture production

Marieke Hoetjes^a, Ingrid Masson-Carro^b

^aCentre for Language Studies, Radboud University, Nijmegen, The Netherlands

^bTilburg center for Cognition and Communication, Tilburg University, The Netherlands

^aP.O. Box 9103, 6500 HD Nijmegen, The Netherlands

^bWarandelaan 2, PO Box 90153, 5000 LE Tilburg, The Netherlands

M.Hoetjes@let.ru.nl, I.MassonCarro@uvt.nl

Abstract

It has been hypothesized that speech and gesture work together, with people relying more on gesture when speaking is hard, and relying more on speech when gesturing is hard. Indeed, previous work showed that when speaking is hard, gesture production can reduce cognitive load and thereby help the speech process. However, it is yet unknown what happens in speech and gesture production when gesturing is hard. In the current study, participants described complex tangram figures. Difficulty in either speech or gesture production was manipulated by having speakers do a secondary task which placed them under either verbal or motoric cognitive load. Results showed an increase in representational gesture rate when participants were under motoric load, as compared to the baseline condition, but only a weak effect of verbal load, and no difference between the verbal and motoric load conditions. Based on these findings, we conclude that making speaking hard caused a marginal increase in gesture production, and making gesturing hard actually led to more gestures. In sum, we find no evidence to support a two-way trade-off between speech and gestures. To our knowledge, this is the first study assessing the effects of a secondary motoric task on gesture production.

1 Introduction

Human communication is multimodal, with speakers typically using both speech and gesture to express an intended meaning. Gesture and speech tend to be temporally and semantically co-expressive (Kendon, 2004; McNeill, 1992). This close relationship between speech and gesture is apparent, for example in the gesture production by congenitally blind speakers who have never seen someone gesture (Iverson & Goldin-Meadow, 1998), in the parallel development of speech and gesture production in children (Gullberg, De Bot, & Volterra, 2008), and in the parallel breakdown in cases of disfluency (Seyfeddinipur, 2006). Although the close relationship between speech and gesture is undisputed, the question of exactly why we gesture when we speak is still under discussion. One view is that (some) gestures are produced to help the listener (Alibali, Heath, & Myers, 2001). Another, not mutually exclusive, suggestion is that producing gestures is done for the speaker herself, as it helps reduce cognitive load, which is needed for speech production.

Several studies have addressed the possible relation between gestures and cognitive load. For example, Goldin-Meadow, Nusbaum, Kelly and Wagner (2001) asked participants to explain how to solve math problems while doing a mentally taxing secondary task (remembering lists of words or letters). They showed that the participants who were allowed to gesture while they were explaining the math problem were better at the secondary task, suggesting that producing gestures frees up mental resources. Melinger and Kita (2007) also gave their participants a secondary task, in their case while describing spatial pictures they had previously memorized. However, Melinger and Kita used two types of secondary tasks: one spatial task for which the same resources as in the primary task had to be used, and one non-spatial task for which other mental resources are required. Speakers produced more gestures during picture description when they had to do a spatial

secondary task compared to a non-spatial secondary task. Again, this suggests that gesture production can help reduce cognitive load, in particular when the secondary task requires the same mental resources as the primary task.

The idea that producing gesture can offload mental space required for speaking has also been suggested by De Ruiter, Bangerter and Dings (2012) in their trade-off hypothesis, which states that when speaking becomes harder, speakers will rely more on gestures, and when gesturing becomes harder, speakers will rely more on speech. De Ruiter and colleagues tested one part of the trade-off hypothesis, namely whether speakers rely more on gesture when speaking becomes harder. They conducted a study in which participants had to describe simple and complex tangram figures repeatedly, and found that gesture rate was not higher for the initial descriptions and complex figures (i.e. when speaking was hard) as compared to the repeated descriptions and simple figures (i.e. when speaking was relatively easy). Instead, gesture production mirrored speech production across experimental conditions, thereby supporting the hypothesis that gestures and speech go hand in hand. Their study, however, targeted the difficulty of verbal referring separate from memory resources—which were kept constant across experimental conditions by having the speaker be able to see the stimuli throughout the task. Given the previously found effects of memory load on gestures (e.g., Melinger & Kita, 2007), it is likely that we would find an effect on the amount of gestures produced when verbal memory resources are taxed, which is another way of making speaking harder. Furthermore, De Ruiter and colleagues did not test the second part of their hypothesis which proposes that speakers will rely more on speech when gesturing gets harder.

In the present study, we designed an experiment to test both ends of the abovementioned trade-off hypothesis. That is, we will make either speaking harder or gesturing harder, by putting participants under one of two types of cognitive load: verbal memory load or motoric memory load. Verbal load will make speaking harder, and, assuming that gestures are generated from action processes (e.g. Kita & Özyürek, 2003) and require spatio-motoric working memory, motoric load will make gesturing harder. As in Goldin-Meadow, et al. (2001) and in Melinger and Kita (2007), cognitive load is imposed in a secondary working memory task. The primary task is for participants to describe complex spatial figures. Describing complex spatial figures requires both verbal and motoric resources. This means that both types of secondary tasks require part of the mental resources also needed for the primary task. The hypothesis is that in cases of verbal load, speakers will rely more on gesture (and thus produce more gestures in the primary task), and in cases of motoric load, speakers will rely more on speech (and thus produce fewer gestures in the primary task), relative to a baseline condition without a concurrent secondary task.

2 Methods

Pairs of participants engaged in a director-matcher task. In the first part of the experiment, one participant (the director) described a series of tangram figures to her interlocutor (the matcher), who had to locate and mark these objects on a visual grid. In the second part, participants exchanged roles, with the interlocutor becoming the director and describing a new set of figures. Each pair of participants accomplished the task under one of three conditions: verbal load, motoric load, or baseline.

2.1 Participants

72 students from Tilburg University—46 female and 26 male, $M = 21$ years—took part in the experiment in pairs, in exchange for partial course credit. All participants read and signed an ethics consent form prior to the commencement of the task, and were informed that they could withdraw their participation anytime.

2.2 Stimuli

Two different sets of ten abstract geometric figures were digitally created, inspired by the Chinese game of tangram (see Figure 1). Each set of figures was compiled into a presentation document, where each figure fully occupied one slide. Version A featured figures 1 to 10, and version B featured figures 11 to 20. Each participant was presented with either version A or version B (see Table 1 below).

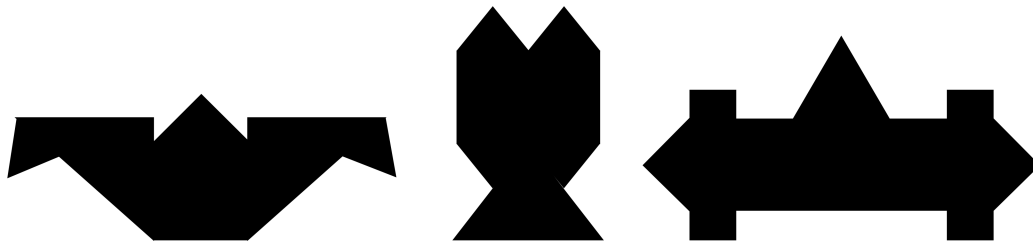


Figure 1. Example of three target tangram stimuli figures.

We induced load in the speakers by means of a concurrent working memory (WM) task that tapped on either verbal or spatio-motoric working memory. In the verbal condition, each tangram figure was preceded by a slide with a word on it (either *dal* or *bal*, the Dutch words for *valley* and *ball*) and followed by a slide instructing the speaker to reproduce the word she had read before. In the motoric condition, each tangram figure was preceded by a short video where an actor performed the British Sign Language sign for either “green” or “brown” (two visually similar signs for hearing speakers with no sign language knowledge) and followed by a slide instructing the speaker to reproduce the movement she had seen before. In the baseline condition, all slides consisted only of tangram figures (see Figure 2).

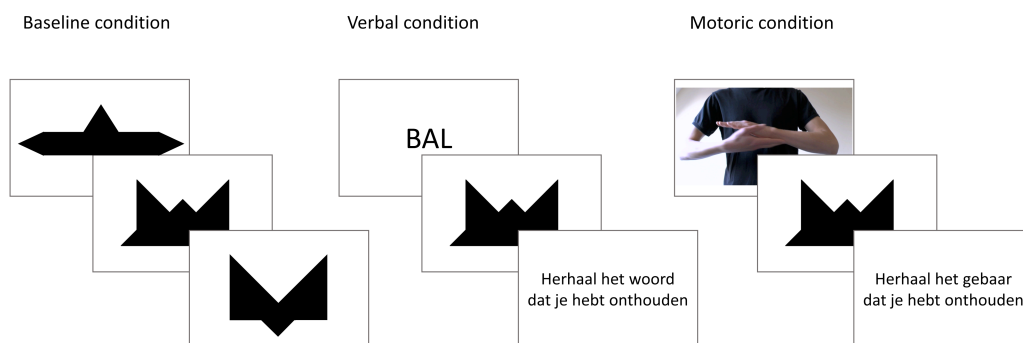


Figure 2. Summary of the three experimental conditions: baseline, verbal, and motoric. The baseline condition shows 3 example trials, and the verbal and motoric load conditions show one example trial with their respective WM tasks.

2.3 Procedure

Each pair of participants was assigned to one of the experimental conditions: verbal, motoric, or baseline (see Table 1). Participants were assigned the roles of director and matcher, and sat at opposite sides of a table. The setup was arranged in such way that there was visual contact between director and matcher, but the matcher could not see the director's screen displaying the task, and the director could not see the matcher's visual grid with all the numbered tangrams. A camera recorded the director's speech and upper bodily movements.

The experiment started with two practice trials, followed by ten target trials. Each trial corresponded to the description of one tangram figure. The director's task was to describe to the matcher each tangram figure displayed on the laptop screen. The matcher's task was to find the tangram figure that was being described on a printed visual grid displaying 12 tangrams, and to write down its corresponding number on an answer sheet. During the description phase, the experimenter remotely controlled the presentation, allowing speakers to describe the figures at their own pace, and moving to the next trial only when the matcher finished writing his answer down. Conversation was not restricted between director and matcher, but it was not encouraged either, and the use of gestures was not mentioned.

In the baseline condition, all ten target figures were consecutively described. In the load conditions, a stimulus (either a word or a video showing a series of movements) was presented prior to each tangram figure, remaining on screen for 3 seconds. The director was instructed to look at the stimulus and memorize it, as she would have to reproduce it (orally, or motorically) after describing the tangram figure (see Figure 2).

After the director had described the ten target figures, participants switched roles, and the previous matcher became the director, describing a new set of ten figures under the same experimental condition.

Table 1. Administration of the experimental conditions

PAIR	SPEAKER 1	SPEAKER 2
Pair 1	Baseline version A	Baseline version B
Pair 2	Verbal version A	Verbal version B
Pair 3	Motoric version A	Motoric version B
Pair 4	Baseline version B	Baseline version A
etc...

2.4 Data annotation

Speech was transcribed verbatim per trial by a native Dutch speaker, using the multimodal annotation tool ELAN (Sloetjes & Wittenburg, 2008). All hand gestures accompanying speech during the description of the tangram figures were identified and classified by two independent coders as representational or non-representational (McNeill, 1992). Adaptors (e.g., touching one's hair) and other irregular movements were excluded from the annotations. The representational, and non-representational gesture rates (number of gestures in proportion to 100 words spoken) were computed. Both coders annotated videos from all conditions, with an overlap of ten videos used to check for the inter-rater reliability of gesture identification. Cohen's κ revealed substantial agreement between coders with respect to the number of gestures produced by speakers ($\kappa = .84$, $p < .001$).

2.5 Statistical analyses

We used linear mixed models (Barr, Levy, Scheepers, & Tily, 2013) to analyze the effects of cognitive load (verbal, motoric, baseline), on our dependent variables (representational and non-representational gesture rate, and the number of words spoken). Participants and items (tangram figures) were included as random factors in the analyses.

3 Results

The task generated a total of 2163 representational and 209 non-representational gestures. We eliminated from our analyses data from 9 participants (three per condition) who produced less than one gesture in total. This resulted in the analysis of 2162 representational and 209 non-representational gestures.

We found an effect of motoric load on the representational gesture rate ($\beta = 2.98$, $SE = 1.25$, $p = .02$), indicating that speakers in the motoric load condition ($M = 8.61$, $SD = 6.18$) gestured more than speakers in the baseline condition ($M = 5.58$, $SD = 5.04$) (see Fig. 3). A similar effect was observed for speakers under verbal load ($M = 7.82$, $SD = 6.1$), albeit statistically weak ($\beta = 2.21$, $SE = 1.22$, $p = .07$). No differences were found between speakers in the motoric and verbal load conditions, as can be seen in Figure 3. No effects of either motoric ($M = .87$, $SD = 2.7$) ($\beta = .11$, $SE = .28$, $p = .69$) or verbal load ($M = .64$, $SD = 1.79$) ($\beta = .06$, $SE = .27$, $p = .81$) were found for non-representational gesture rate, in comparison with the baseline condition ($M = .75$, $SD = 1.72$).

Moreover, there was an effect of motoric load on the average number of words spoken (see Figure 3). Participants in the motoric load condition ($M = 35.62$, $SD = 18.44$) used fewer words on average than participants in the baseline condition ($M = 45.87$, $SD = 31.14$) ($\beta = -10.11$, $SE = 4.12$, $p = .01$). No effects of verbal load were found ($M = 44.12$, $SD = 25.57$) ($\beta = -1.18$, $SE = 4.02$, $p = .76$).

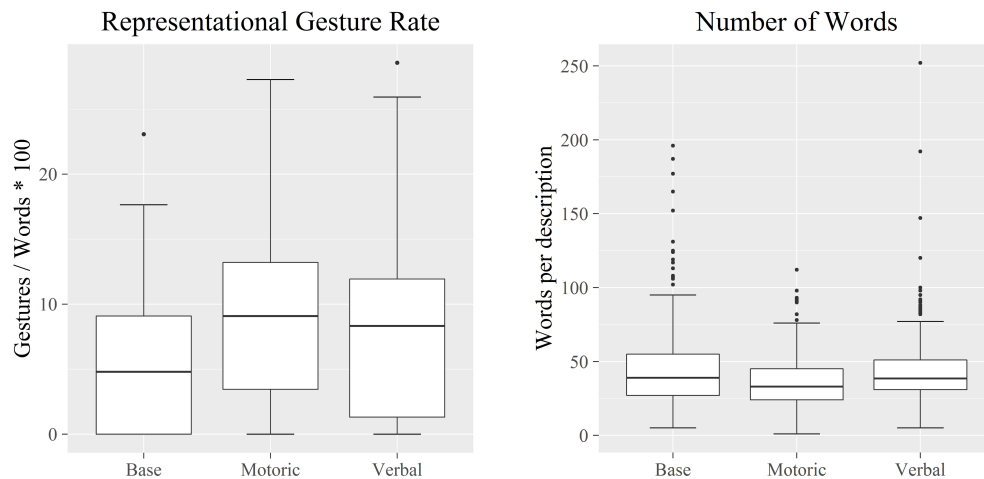


Figure 3. Box plots of the speaker's gesture rate (left) and number of words per description (right), for each of the experimental conditions (baseline, motoric load, and verbal load).

4 Discussion

This study set out to investigate the effect of verbal and motoric cognitive load on gesture production. The aim was to address the question whether gesture production helps reduce cognitive

load, and in particular whether, as the trade-off hypothesis (de Ruiter, et al., 2012) suggests, people rely more on gesture when speaking is hard, and rely more on speech when gesturing is hard. We created difficulty in speaking or in gesturing by having participants conduct a secondary task, concurrent to the primary task of describing complex spatial tangram figures. In the secondary task, participants had to either remember one of two similar words (verbal load), or one of two similar hand configurations (motoric load). The hypothesis was that speakers would produce more gestures in the primary task if they were under verbal cognitive load—compared to the baseline condition—and fewer gestures in the primary task if they were under motoric cognitive load.

The results showed an increase in representational gesture production when participants were under load, relative to the baseline condition. This effect was prominent for motoric load, but weak for verbal load. Importantly, there was no difference between the two load conditions. Although these findings marginally support one end of the trade-off hypothesis (namely that speakers produce more gestures when under verbal load), they do not support the assumption that tapping into different types of working memory processes causes differences in gesture production. It seems that being under load, because of having to carry out a secondary task, generally led to more gestures.

The fact that there was no difference in gesture production between the verbal and the motoric load conditions can be viewed in several ways. Firstly, it might be the case that, although being under load affects gestures, the type of cognitive load simply does not matter for gesture production. Another possibility is that our manipulation of cognitive load was not successful, in the sense that the secondary tasks participants had to do did not actually tap into *different* types of cognitive load. After all, it can be argued that having to remember words shown on a screen does not have to be a verbal task per se, but could also be a more general memory task. Therefore, in future studies, the material used to induce verbal load could be presented orally instead of visually. A third possibility to explain the findings is that both our verbal and motoric load conditions led to more gesturing *for different reasons*. That is, it could be that having speakers keep words in mind while producing verbal descriptions pushed them to rely more on their hands, but that keeping hand configurations in mind while talking simply “boosted” gesture production. This would be compatible with simulation-based accounts of gesture production (e.g., Hostetter & Alibali, 2008), which posit that representational gestures arise from perceptual and motoric simulations underlying thinking and speaking. In this regard, it is possible that maintaining a movement sequence active while performing a side task just increased the amount of motor activation experienced, leading to an increase in the production of overt gestures.

A point of discussion could be whether both types of secondary tasks were equally difficult for the speakers. The tasks were chosen because they are both deceptively easy; although the words and signs themselves were simple, the similarity between the words or signs made the task quite hard. One possible way to study this would be by looking at the accuracy of the answers given in the secondary tasks. The question then is, though, what information we can glean from this. Would an incorrect answer in the secondary task mean that the speaker was not paying attention, or that the speaker was paying attention but the task was harder?

It’s important to note that the results found applied only to representational gestures, and were not mirrored by either non-representational gestures, or speech. For speech, we found that fewer words were used when participants were under motoric load, compared to the baseline condition. It is likely that the motoric task more strongly activated spatio-motoric thinking, resulting in more information expressed through hand gestures and less information expressed through speech. Further analyses on the semantic content of the speech and gestures produced are needed to test this idea.

In conclusion, the present study has provided additional evidence, in line with Goldin-Meadow, et al. (2001) and Melinger and Kita (2007) that speakers produce more gestures when their memory

resources are taxed. Furthermore, we proposed the use of a novel task to study spatio-motoric resources during multimodal language production.

Acknowledgements

The research reported in this article was financially supported by The Netherlands Organisation for Scientific Research (NWO) [grant 322-89-010]. We thank Ivo de Ruiter and Chiara de Jong for their invaluable help in collecting and annotating parts of the video dataset.

References

- Alibali, M., Heath, D. C., & Myers, H. J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language*, 44, 169–188.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255–278.
- de Ruiter, J. P., Bangerter, A., & Dings, P. (2012). The interplay between gesture and speech in the production of referring expressions: Investigating the tradeoff hypothesis. *Topics in Cognitive Science*, 4(2), 232–248.
- Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, 12, 516–522.
- Gullberg, M., De Bot, K., & Volterra, V. (2008). Gestures and some key issues in the study of language development. *Gesture*, 8(2), 149–179.
- Hostetter, A. B., & Alibali, M. W. (2008). Visible embodiment: Gestures as simulated action. *Psychonomic Bulletin and Review*, 15, 495–514.
- Iverson, J., M., & Goldin-Meadow, S. (1998). Why people gesture when they speak. *Nature*, 396, 228.
- Kendon, A. (2004). *Gesture. Visible action as utterance*. Cambridge: Cambridge University Press.
- Kita, S., & Özyürek, A. (2003). What does cross-linguistic variation in semantic coordination of speech and gesture reveal?: Evidence for an interface representation of spatial thinking and speaking. *Journal of Memory and Language*, 48, 16–32.
- McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. Chicago: University of Chicago Press.
- Melinger, A., & Kita, S. (2007). Conceptualisation load triggers gesture production. *Language and Cognitive Processes*, 22(4), 473–500.
- Seyfeddinipur, M. (2006). *Disfluency: Interrupting speech and gesture*. Radboud University Nijmegen.
- Sloetjes, H., & Wittenburg, P. (2008). Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th International Conference on Language Resources and Evaluation (LREC 2008)*.